11

US Army Corps
of Engineers
Hydrologic Engineering Center

GENERALIZED COMPUTER PROGRAM

DATCHK & DATVUE

Data Screening Software

User's Manual

February 1995

CPD-65

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution of this document is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CPD-65	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION USA Corps of Engineers Hydrologic Engineering Center 6b. OFFICE SYMBOL *(If applicable)* CEWRC-HEC	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS *(City, State, and ZIP Code)* 609 Second Street Davis, California 95616	7b. ADDRESS *(City, State, and ZIP Code)*
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS *(City, State, and ZIP Code)*	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*

DATCHK & DATVUE - Data Screening Software - User's Manual

12. PERSONAL AUTHOR(S)

CEWRC-HEC

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Computer Program	FROM ______ TO ______	February 1995	65

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	Screening of hydrologic data

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The software described in this User's Manual is intended for screening of hydrologic data. The software can function "on-line" for screening of real-time data, or it can be used to screen historical data. The software is intended for use on UNIX-based workstations and requires use of the Hydrologic Engineering Center Data Storage System (HEC-DSS).

Computer program DATCHK performs screening tests with user-specified criteria and classifies each data value as missing, erroneous, questionable or o.k.. Six types of test are available: absolute magnitude, duration magnitude, rate-of-change, constant value, relative magnitude and distribution. Replacement values for flagged data can be generated automatically.

Computer program DATVUE is an interactive program that produces graphical and tabular data displays and provides the capability to rapidly edit data previously screen by DATCHK.

This User's Manual describes capabilities of DATCHK and DATVUE, and specifies input requirements. Examples of program usage are also provided.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL John C. Peters	22b. TELEPHONE *(Include Area Code)* (916) 756-1104 22c. OFFICE SYMBOL CEWRC-HEC

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

DATCHK & DATVUE

Data Screening Software

User's Manual

February 1995

US Army Corps of Engineers
Hydrologic Engineering Center
609 Second Street
Davis, CA 95616-4687
(916) 756-1104

CPD-65

Table of Contents

Appendices

List of Figures

Figure **Page**

Preface

The software described in this Manual is intended for error-screening of hydrologic and other environmental data. The software can be used on UNIX-based workstations or Harris minicomputers, and is implemented in conjunction with the HEC Data Storage System. The data to be screened may be "on-line" (i.e., incoming from a data network), or historical.

The software was designed by the following HEC personnel: Art Pabst, John Peters, and William Charley; and by Paul Ely, consultant. The software was written in FORTRAN by Paul Ely. This manual was prepared by John Peters and Paul Ely.

Chapter 1

Introduction

A common problem in dealing with hydrologic and other environmental data is that data values may be erroneous due to problems in measurement, transmission, processing, etc. If such data are being used in "real-time", it is generally necessary to perform on-line data screening to preclude the use of faulty information in the making of forecasts and operational decisions.

The software described herein can be used to apply screening tests with user-specified criteria to flag (categorize) each data value as follows:

- value is <u>missing</u>
- value should be <u>rejected</u>; the value failed one or more severe (i.e., reject-level) tests
- value is <u>questionable</u>; the value failed one or more questionable-level tests
- value is <u>o.k.</u>; all tests were passed

Once data is so classified, a manual review can focus on just that subset of data that falls in the first three categories - missing, rejected or questionable. The software facilitates such manual review by providing graphical and tabular data displays and the capability to rapidly process flagged values with the following choices:

- accept the original data value
- accept a software-generated replacement value
- replace the original data value with an interpolated or user-specified value
- classify the data value as missing

At present there is capability to perform six types of tests: absolute magnitude, duration magnitude, rate-of-change, constant value, relative magnitude and distribution. The first four tests are performed using data only from the site of interest, whereas the relative magnitude and distribution tests require data from nearby sites. The tests are described in Chapter 2, and associated input requirements in Chapter 3.

Two sets of criteria can be specified for a test: a set of stringent criteria to serve as a basis for rejecting data, and a set of less restrictive criteria to flag questionable data. A tradeoff is invariably involved in setting such criteria. If criteria are relatively rigorous, there will be a tendency to flag too many 'good' data values. Conversely, less restrictive criteria will tend to result in too many 'bad' data values not being flagged.

Application of the software requires use of the Hydrologic Engineering Center Data Storage System (HEC-DSS). That is, it is assumed that data has been entered into a DSS file from a real-time data acquisition system, or by loading historical data from an external source (e.g., CD-ROM) via a DSS utility program.

Two programs are employed for data screening:

1. DATCHK performs screening tests with user-specified criteria and optionally generates replacement values for flagged data.

2. DATVUE displays data flagged by DATCHK and allows the user to edit replacement values.

DATCHK reads raw (unscreened) data from a DSS file, applies the requested tests, and writes revised data to a DSS file. The revised data include a flag that describes the quality of the raw data, why it is questioned, and how the value was revised. DATCHK can also maintain a data-status table showing the quality category (o.k., reject, questionable, missing) associated with the data. In a real-time environment, DATCHK would generally be executed automatically (on-line) as data is received.

DATVUE is an interactive program that displays graphs and tables of raw and revised data for visual review. Revised data can be replaced with interpolated values, or revised data can be edited manually. Values changed by the user in DATVUE are 'protected' so that they will not be subsequently changed by DATCHK.

DATVUE performs several additional functions. DATVUE converts test criteria from a text file to a random-access (DSS) file for use by DATCHK. DATVUE is used to view the log file written by DATCHK, edit a list of faulty gages, create a data-status table to be updated by DATCHK, and view the data-status table. Chapter 4 describes capabilities and usage of DATVUE.

The current versions of DATCHK and DATVUE are intended for application on UNIX-based workstations and HARRIS minicomputers. The computer program DSPLAY is required if it is desired to have DATVUE show plots of data in addition to tabular displays.

Application of the data-screening software requires initial development of a number of files and macros. Chapter 5 summarizes information on file usage and provides guidelines and illustrations for implementing screening procedures.

(This page intentionally left blank)

Chapter 2

Screening Program (DATCHK)

2.1 Introduction

This Chapter describes the screening tests that can be performed by DATCHK, and discusses the role of associated input and output files. Input requirements for DATCHK are specified in Chapter 3.

2.2 Overview

Data screening is performed within a user-specified time window. For example, for real-time applications, the time window might be specified as ending at the current time and beginning a specified period prior to the current time. This time window is referred to as the update time window.

In order to perform some tests (e.g., the constant value test) on data at the beginning of the update time window, it is necessary to have access to data that occurs prior to this window. Hence a context time window can also be specified to enable data to be screened in relation to data outside the update time window.

Time-series data (at a regular or irregular time interval) is read from a raw-data DSS file. Data which has been previously reviewed and 'protected' by DATVUE is merged with the raw data, and the combined data is screened.

Data is screened using tests described subsequently in this Chapter. If requested, a replacement value is estimated for each questionable or missing value in the update time window. An entry to a log file is written for each questionable value.

Screened data, including replacement values and data-quality flags, are written to a DSS file for the update time window. Figure 2.1 illustrates data flow associated with DATCHK/DATVUE.

2.3 Preliminary Processing

Before applying the screening tests, DATCHK performs the following operations which prepare the data for screening.

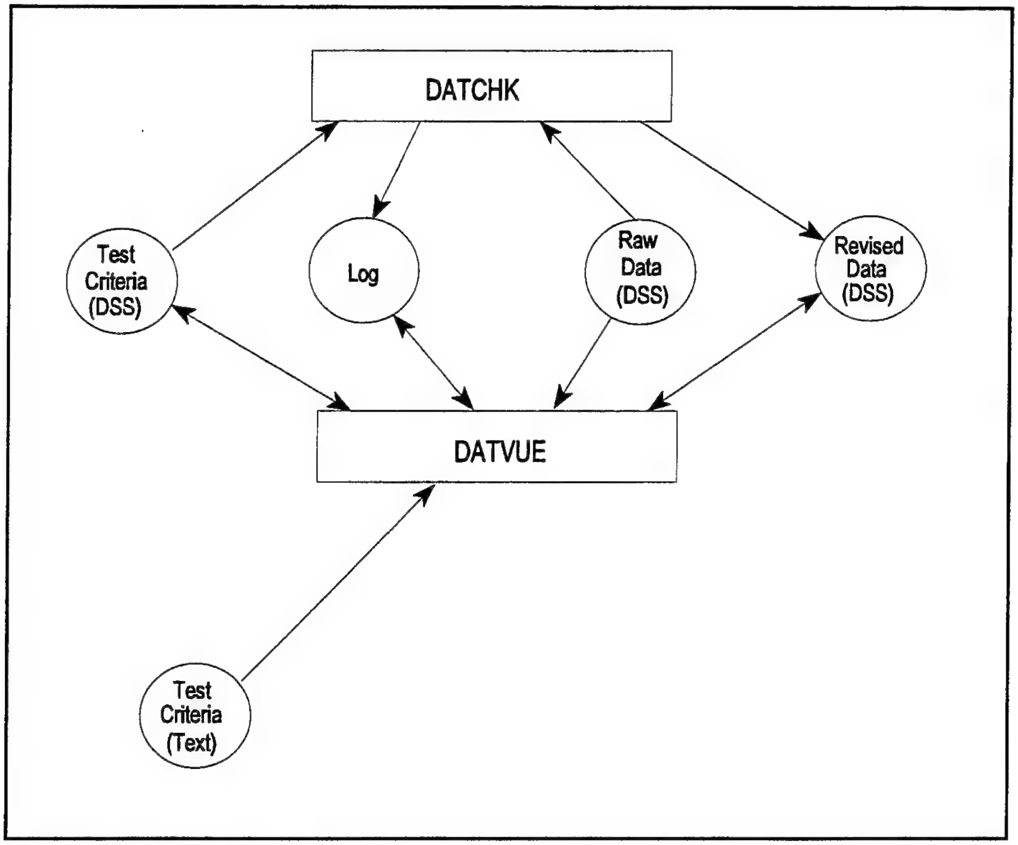


Figure 2.1 Data Flow

DATVUE converts a test criteria (text) file into a DSS file to be used by DATCHK. DATCHK accesses "raw" DSS data and performs the specified tests. DATCHK writes revised data to a "revised data" file and writes information regarding data quality to a log file. The raw-data file, revised-data file and log file can be examined with DATVUE.

2.3.1 Protected Values

Before screening raw data, DATCHK reads revised data from earlier screenings. If a value is protected the raw value is replaced with the protected value, so the protected values are used for screening. This may change some of the screening results. Also the protected value will not be changed by DATCHK when DATCHK estimates replacement values.

Whenever a value is changed by DATVUE, that value is protected. DATCHK merges the protected data with raw data and screens the resulting time series. DATCHK will not change protected data.

2.3.2 Faulty Gage

Sometimes a gage is know to be broken or reporting erroneous data. The gage can be referenced in a list of "gages to be ignored", along with an associated time period. All data for this gage will be rejected during this period.

2.3.3 Incremental Values

Data values of the instantaneous cumulative type ("INST-CUM", in DSS terminology), such as cumulative precipitation, are converted to incremental values before screening tests are applied. When a negative incremental value occurs, DATCHK automatically checks the two cumulative values involved against a straight line passing through cumulative data values in the vicinity of the negative increment. The cumulative value farthest from the line is rejected, and the incremental value is treated as missing.

When cumulative data "steps down", as when a gage is reset to zero, the two values producing the negative step are labeled "questionable", and the negative incremental value is set to zero.

2.4 Screening Tests

Six tests are described below. Two sets of criteria can be specified for each test. One set can be used to reject data and a less stringent set to flag questionable data. Once a value has been rejected, it is marked as missing and will not be used in subsequent tests.

2.4.1 Absolute Magnitude

This test simply determines whether a data value falls within user-specified upper and lower limits. A value falling outside the range is flagged. For stage data, a minimum value at the reject level would typically be the stage at the bottom of the staff gage; a minimum value at the questionable level might be a 10%-chance (10-year) low-flow stage. An upper limit for rejecting a stage might be an estimate for the probable maximum stage, whereas at the questionable level it might be a 10%-chance high-flow stage.

2.4.2 Duration Magnitude

This test is essentially the same as the Absolute Magnitude test, except that values are accumulated over a specified duration, and the accumulated value is compared against the specified limits. The test can be specified for each of 5 durations, for each level (reject and questionable).

For example, for precipitation data, the lower limit is normally zero. The upper limits for questionable data might be the 10-year rainfall for 1-hour, 3-hour, 6-hour, 12-hour and 24-hour periods. The upper limits for rejecting data values would be the probable maximum precipitation for the same periods.

2.4.3 Constant Value

Data values are checked for being constant over a specified duration. If values are constant for a duration equal to or greater than the specified duration, all values for that duration are flagged. This test can be run with the condition that the constant value is greater than a specified minimum value. For example, it would generally be desirable to not flag a continuous string of "0" incremental precipitation values. It is also possible to specify a tolerance such that the "constant" value can vary within the tolerance and still be considered constant.

2.4.4 Rate-of-Change

Data values are checked to determine that the change in magnitude from one value to the next does not exceed positive or negative limits. When a rate-of-change limit is exceeded, the two associated values are checked to determine which is closer to a linear least squares regression line passing through the four data values preceding the pair and the four values following the pair. The value which is farther from the regression line is flagged.

2.4.5 Relative Magnitude

Data is checked to determine whether it is within a range about "base" data. Data outside that range is flagged. Base data can be generated from one or more similar time series measured at other locations. A moving average or routing transform can also be applied to base data. For

example, suppose that discharge values for gage "B" are being screened, and there is another gage, "A", upstream. A test could be specified such that the discharge values for gage "A" are routed to gage "B", and gage "B" values are flagged if they do not fall within a specified range of the routed values.

2.4.6 Distribution

A primary impetus for creation of this test was the need to detect data from blocked or malfunctioning precipitation gages. The test compares the distribution of incremental data values for the test gage against distributions for two other "base" gages. Time series for the three gages are partitioned into events consisting of periods of simultaneous occurrence of zero and non-zero values. Values reported for an event are placed into three classes. The class distribution for the test gage is compared against distributions for each base gage using a chi-squared variable. If the probability of the chi-squared variable is less than a user-specified threshold probability, all values in the event are flagged. An illustration of this test is provided in Appendix B.

2.5 Estimated Values

Two options exist for replacing rejected data values: the values can be set to "missing", or values can be generated by linear interpolation between adjacent valid data values (i.e., raw data not rejected or questioned).

Three options exist for questionable data: the raw data values can be accepted as valid, replacement values can be interpolated, or the values can be set to "missing".

Missing raw data values can also be replaced by interpolation.

When interpolation is selected as a means for generating replacement values, a limit is specified for the maximum number of contiguous values to be interpolated. If this limit is exceeded, interpolation is not performed, and replacement values are set to "missing". Interpolation is based on raw data in conjunction with data previously protected with DATVUE.

2.6 Input Files

DATCHK reads test criteria from a DSS (random access) file. The user first creates a text file containing the test criteria, then generates a DSS file from the text file using DATVUE. The execution line parameter CTEXT is used to identify the text file when DATVUE is run in batch mode. The CONVERT command can be used when DATVUE is run in interactive mode. DATVUE reads data from the text file and writes the test criteria to the DSS file. One record is written for each location and parameter. The pathname used for this record is the same as the pathname of data to be screened, with the data (part D) replaced by /-SCREEN-/. (See Chapter 5.)

DATCHK screens time series for each record listed in the catalog of the DSS criteria file. An optional input file, listing selected pathnames, can be used to override this automatic screening process. This file is named using the execution-line parameter INPUT.

DATCHK can also read a gage-list file created by DATVUE. The gage list is a random-access file listing gages known to be defective. Data reported by these gages for specified time windows is rejected.

DATCHK can maintain a data status file that summarizes data quality information, and optionally categorizes data values in magnitude ranges. The data status file is a random access file created by DATVUE from a list of pathnames. The data status file can be displayed with DATVUE.

2.7 Seasonal Variation of Test Criteria

It may be appropriate to vary criteria for some tests seasonally or to reflect other changing conditions. Separate criteria (text) files can be developed for each season. Corresponding DSS criteria files can then be generated, or a single file can be changed seasonally using the appropriate text file.

2.8 Output

DATCHK writes information to several files:

1) Screened data with flags are written to the destination DSS file. Data written to the destination file may have been revised by DATCHK (i.e., replacement values may have been generated). If the data is not revised, it is still copied to the destination file. DATVUE can be used to edit and modify the destination-file data.

2) A log file is maintained that contains one line for each data value when a revised value is different from the raw data. Job-progress messages, error messages, and an end-of-job summary are also written to this file. The log file is a binary file named "Lyymmdd", corresponding to the year, month and day when the message is written. DATVUE can be used to view the contents of the log file.

3) A data-status file can be maintained by DATCHK. The file is a shared-access binary file originally created by DATVUE. The data-status file shows the quality of data for the last 44 hours of data screened by DATCHK. DATVUE can be used to view the contents of the data-status file.

4) Error messages are written to an output file. This file is specified using the OUTPUT parameter on the execution line.

Chapter 3

Specification of Test Criteria

3.1 Overview

This Chapter describes input requirements for screening tests. The criteria are first defined in a text file. A DSS criteria file (for use by DATCHK) is generated from the text file with DATVUE.

Each instruction in the (text) criteria file begins with a key word, such as DEFINE, END, ASSIGN, COMPUTE, TIME, DATA, etc. Lines beginning with an asterisk (*) and blank lines are ignored. Hence comments can be inserted in the file. An instruction beginning with the key word DATA specifies the pathname associated with a location and variable that is being screened. When such an instruction is encountered (by DATVUE), previously specified criteria for that location/variable are processed and a specially formated record is written to the DSS criteria file.

A test delimited with DEFINE and END instructions is global in nature. That is, it is given a label and can be invoked (with a TEST instruction) for any subsequent DATA records. A test can also be specified on an ad hoc basis, in which case it only applies until criteria are reset (as described below). The following is an illustration of an instruction sequence.

Tests are defined at the beginning of the file:

```
DEFINE TEST1
.
.
.
END        (end of TEST1)
DEFINE TEST2
.
.
.
END        (end of TEST2)
```

Then a test is named and time series are listed for that test:

```
TIME      (Specify a time window)
TEST  TEST1
DATA      (data to be screened using TEST1)
DATA
* Use another test
TEST  TEST2
DATA      (data to be screened using TEST2)
DATA
```

Set criteria to be used immediately for following data

```
CRITERIA
DATA   (data to be screened with ad hoc criteria)
DATA
END    (end of file)
```

3.2 Instruction Summary

The following is a list of available commands:

Command	**Function**
*	comment
ALARM	set limits which trigger an alarm
ASSIGN	assign a label to a pathname
COMPUTE	assign a label to a time series generated as a function of one or more time series
CONTEXT	set context time window
CRITERIA	specify test criteria
CRITFILE	set name of DSS file to contain test criteria
DATA	file and pathname for raw and revised data
DEFINE	begin test definition
END	end test definition or instruction file
ESTIMATE	set estimation method for revised data
PRECISION	set number of decimal places for DATVUE display
RANGE	set range limits for data-status report
TEST	identify test to use for following data
TIME	set update time window

The instructions ALARM, ASSIGN, COMPUTE, CRITERIA, ESTIMATE, PRECISION, and RANGE can be used to define a global test delimited with DEFINE and END

The instructions ALARM, ASSIGN, COMPUTE, CRITERIA, ESTIMATE, PRECISION, and RANGE can be used to define a global test delimited with DEFINE and END instructions, or to define an ad hoc test.

A TEST instruction lists previously-defined (global) tests which will be applied to time series identified on subsequent DATA instructions.

A test applies to all DATA instructions until the test criteria are reset. Test criteria are reset when an ASSIGN, COMPUTE, CRITERIA, DEFINE, ESTIMATE, PRECISION, RANGE, or TEST instruction is read after a DATA instruction.

CRITFILE and TIME instructions remain in effect until changed.

CONTEXT parameters are reset to zero when TIME is read. The context time window will be the same as the update time window when a TIME instruction is read, so if a larger context window is desired a new CONTEXT instruction is required when TIME is redefined. If CONTEXT is included in a test definition, the context time window will be defined when that test is used.

Key words may be abbreviated to two or more characters as long as the abbreviation uniquely identifies a key word. For example, CRITFILE can be abbreviated to CRITF.

Angle brackets, < >, are used to delineate a description of an item. Items enclosed in square brackets, [], are optional. Where one of several choices are to be used, the choices are separated using a pipe |, e.g. + | - means choose + or -.

<label> is used as a shorthand notation for a pathname. Labels consist of one to eight alphanumeric characters, A-Z and 0-9; the first character must be alphabetic, A-Z. Labels are assigned using the ASSIGN or COMPUTE instructions. The label 'DATA' is automatically assigned to the time series currently being screened.

Wherever <pathname> occurs in this input description the following structures may be used.

filename: /a/b/c/d/e/f/
filename: A=a B=b C=c D=d E=e F=f
/a/b/c/d/e/f/
A=a B=b C=c D=d E=e F=f

where a, b, c, d, e, f are appropriate pathname parts or blank.

Pathname part D may be specified, but it will be ignored and will be determined from the TIME specification.

On Unix systems file names area case-sensitive, so lower case characters are retained for file names. All other instructions are converted to upper case characters.

3.3 Example

This section shows an example of a text file for setting test criteria. The commands are described in section 3.4. Chapter 5 contains additional examples.

```
* Sample test criteria
* Comments begin with an asterisk
*
* Set file to hold test criteria for DATCHK
CRITFILE CRITDSS
*
* Set time window for current data (last 3 days)
TIME T-3D T
* Or
* Set time window for historical data
* Use TIME 23MAR1989 2400 30MAR1989 2400
*
* Set context time window
CONTEXT START - 24H
*
*
* Define duration-magnitude test for precipitation

DEFINE P _MAG
   CRITERIA DUR Q 0 1.7   1H
   CRITERIA DUR Q 0 2.4   3H
   CRITERIA DUR Q 0 2.8   6H
   CRITERIA DUR Q 0 3.1 12H
   CRITERIA DUR Q 0 3.3 24H
*
   CRITERIA DUR R 0 14.4   1H
   CRITERIA DUR R 0 20.3   3H
   CRITERIA DUR R 0 25.9   6H
   CRITERIA DUR R 0 35.5 12H
   CRITERIA DUR R 0 41.5 24H
END
*
* Apply tests to precipitation data
* Replace reject data with missing values
* Write revisions to file REVDSS
```

```
   TEST P _ MAG
   ESTIMATE R MISSING
   DATA RAWDSS:/SMITH/ABC/PRECIP-INC//1HOUR/OBS/; REVDSS:
   DATA RAWDSS:/SMITH/DEF/PRECIP-INC//1HOUR/OBS/; REVDSS:

*Add distribution test for gage JKL

   TEST P_MAG
   ASSIGN A=RAWDSS:/SMITH/ABC/PRECIP-INC//1HOUR/OBS/
   ASSIGN B=RAWDSS:/SMITH/DEF/PRECIP-INC//1HOUR/OBS/
   ESTIMATE R MISSING
   CRITERIA DIS Q 6H .02 A B
   DATA RAWDSS:/SMITH/JKL/PRECIP-INC//1HOUR/OBS/; REVDSS:
*
* Test flow data
* Replace questionable/rejected flow values by linear interpolation for up to three
*    contiguous values

   CRITERIA ABS Q 1200  31000
   CRITERIA ABS R    0 690000
   ESTIMATE QR LINEAR 3 ACCEPT
   DATA RAWDSS:/SMITH/MNO/FLOW//1HOUR/OBS/; REVDSS:
END
```

To test the data, first convert the text file (here labeled CRITIN) to a DSS file using DATVUE:

```
datvue ctext=CRITIN
```

Then screen the data using the criteria file, CRITDSS, created by DATVUE:

```
datchk critfile=CRITDSS
```

Note that the name of the DSS criteria file has been set with the CRITFILE instruction in the example input. Following execution of DATCHK, DATVUE can be used to review and edit the screened data.

3.4 Test Criteria Instructions

The following is a detailed description of each criteria-file instruction:

*** <comment>**

May be used to annotate instruction stream. Lines beginning with an asterisk are ignored.

CRITFILE <filename>

Identifies the file where test instructions will be stored for use by DATCHK. <filename> is a DSS file. Instructions from the text file are loaded into DSS for use by DATCHK. DATCHK uses the pathname specified on a DATA instruction to locate the criteria in <filename>.

If the DSS file already exits, it will be eliminated when the CRITFILE instruction is read. Do not use the same file name twice in one set of instructions. Do not include the CRITFILE instruction in a test definition.

TIME <start date> <start time> <end date> <end time>
or
TIME T-nX [T-nX]
or
TIME -nX [-nX]

Specify update time window.

T is current time, i.e. time when this instruction is processed.

n is an integer, and
X is H for hours or D for days.

Form 1: Set time window using specific times. Default times will be set to the reference time.
Form 2: Set time window in terms of current time.
Form 3: Set time window in terms of reference time.

Reference time is set on the DATCHK execution line using DATE and TIME parameters. Default for DATE and TIME are the date and time when DATCHK started. If DATE is set to a different date, default for TIME becomes 2400.

Assume Refdate and Reftime are values set by DATCHK execution-line parameters DATE and TIME, or their defaults; then before evaluating the TIME instruction, the default time window is set to:

Refdate, 0001, Refdate, Reftime

Examples:

TIME 01OCT1989 31OCT1989
Time window is 01OCT1989, 0001, 31OCT1989, Reftime

TIME 01OCT1989 31OCT1989 2400
Time window is 01OCT1989, 0001, 31OCT1989, 2400

TIME -7D
Time window is Refdate-7D, 0001, Refdate, Reftime

Messages are written to the log file for any questionable data found in the update time window. Revised data are written to the DSS destination file for the update time window.

CONTEXT_TIME [START-nX] [END+nX]

Specify context time window. DATCHK will use data from the context window while applying tests for questionable data in the update window. Only data in the update window is written to DSS.

nX is an offset from the START or END of the update time window, where n is an integer and X is M, H, or D for minutes, hours, or days, respectively.

Default context window is the update window. Whenever update window is defined the context window is set to be the same as the update window. So the context offsets must be set after the TIME instruction is processed.

DEFINE <name>

Begin a test definition. Tests are defined by ASSIGN, COMPUTE, CRITERIA, ESTIMATE, PRECISION, and RANGE instructions. This test can be referenced later by <name> on a TEST instruction. <name> can have one to eight characters. A test definition is terminated by another DEFINE instruction or a DATA, END, or TEST instruction. Several CRITERIA instructions can be included in a test.

ALARM <min_val> <max_val> <min_rate> <max_rate>

Set limits which will trigger an alarm. If a value is outside the given limits an alarm code will be written to the log file. Alarm limits are

<min_val> - minimum acceptable value
<max_val) - maximum acceptable value
<min_rate> - minimum acceptable rate-of-change
<max_rate> - maximum acceptable rate-of-change

Note: Although alarms are written to the log file, at the present time no use is made (within DATVUE) of the alarm information. It is intended that software will be developed to read the alarm information and activate desired actions, such as generating messages on a display terminal or initiating automated telephonic messages.

ASSIGN [.option] <label> = <pathname> [<min> <max>]

Assign a <label> to a pathname. ASSIGN is used to define a list of pathnames which are associated with the time series being screened. These pathnames may be used in tests applied to this time series.

Options are:

A Apply absolute magnitude test using limits from CRITERIA instructions for time series named on DATA instruction to time series named on this ASSIGN instruction. Rejected data will be replaced with missing values before the time series is used in DATCHK.

R Apply a range check to time series named on this ASSIGN instruction. Data outside specified limits, <min> and <max>, will be replaced with missing values before the time series is used in DATCHK.

Time series defined on ASSIGN and COMPUTE instructions are used internally by DATCHK. These instructions do not change data stored in DSS.

Up to ten labels can be defined using ASSIGN and COMPUTE instructions.

COMPUTE <label> = <function>

Compute a time series to be used in a screening test. This time series is a function of other time series defined by the ASSIGN instruction

Up to ten labels can be defined using ASSIGN and COMPUTE instructions.

Available functions are:

FMA (label, np) computes a forward moving average of the preceding np values in label. np must be 2 or larger. np values at the beginning of the resulting time series will be missing, so a context window beginning at least np values before the update window should be used. If a value in label is missing, the value is ignored and the average is over one less value. At least 2 values of label must be defined, otherwise the result is missing.

ROUTE (label, nsteps, K, x) routes a hydrograph using the Muskingum method. nsteps is the number of routing steps (subreaches). K and x are the Muskingum routing parameters. K is expressed in hours, and will be divided by nsteps to derive the effective K for each subreach.

SUM (label1,...,label4) adds the named time series together. Two to four time series can be combined.

TSHIFT (label, nX) shifts a time series by nX, where n is an integer, and X is M, H, or D for minutes, hours, or days, respectively. When n is negative the time series is shifted backward in time.

WTDAVG (label1, wt1, ..., label4, wt4) computes a weighted-average time series. Two to four time series can be used in this weighted average.

PRECISION <ndec>

Set number of decimal places to show when displaying time-series data in DATVUE. <ndec> is the number of decimal places. If the precision is not set, DATVUE will show from 0 to 3 digits after the decimal based on the average value of the raw data to be displayed.

RANGE <limit1> <limit2> <limit3>

Specify limits of range categories. <limits> are lower limits for three range categories. Ranges appear as -, 1, 2, or 3 in the data-status table. There is a sample data-status table in Chapter 4.

CRITERIA <type> <qflag> <parameters>

Set screening criteria. <type> is the type of test, <qflag> is the quality attributed to data failing this test, and <parameters> are criteria which are used by the specified test.

<qflag> can be 'Q' for questionable data or 'R' for reject data. Test types and their parameters are:

ABSOLUTE_MAGNITUDE <qflag> <min> <max>
abbr: ABS

Specify limits for absolute magnitude test. Values outside the specified range are assigned quality <qflag>.

DURATION_MAGNITUDE <qflag> <min> <max> <duration>
abbr: DUR

Specify limits for duration-magnitude test. Values outside these limits are flagged with <qflag>.

Values are accumulated for the period of time set by <duration>. Data used to compute the sum are flagged if the sum is outside the specified limits.

CRITERIA DURATION-MAGNITUDE instructions can be repeated with limits for up to 5 durations. <duration> is expressed as nX where n is an integer and X is M, H, or D for minutes, hours, or days.

CONSTANT_VALUE <qflag> <duration> [<min_value>] [<tolerance>] [<nmiss>]
abbr: CONST

Specify duration for the constant value test. If values do not vary by more than <tolerance> for the given duration, they will be flagged with <qflag>. Up to <nmiss> contiguous missing values can be included in the string of constant values. Any values less than or equal to <min_value> will not be flagged. <min_value>, <tolerance>, and <nmiss> are assumed to be zero when they are not specified. <duration> is expressed as nX where n is an integer and X is M, H, or D for minutes, hours, or days.

RATE_CHANGE <qflag> <negative rate> <positive rate>
abbr: RATE

Define rates for rate-of-change test. Rate is expressed in units per hour.

RELATIVE_MAGNITUDE <qflag> <min_expr> <max_expr> [<action> <duration>]

abbr: REL

Define the range about a base time series for use in the relative magnitude test. Range is set by the expressions <min_expr> and <max_expr>. The base time series is defined by an ASSIGN instruction or by a COMPUTE function. An expression is of the form:

<number> * <label> [+ | - <number>]

Note: The multiplier is required, but the additive term is optional, e.g. 1*A -50, 2*A.

The optional parameters <action> and <duration> are used to compare values over a period of time. If action is ACCUM, data is accumulated for the given duration for both the base time series and the time series being tested. The accumulated values are used in the test. If the tested value is outside the range, all the values used to determine the accumulated value are marked with qflag.

When the ACCUM action is used, values are accumulated over a specified period of time and compared with the accumulated values from the base time series over the same period of time. This time period ends at the later of the time for the test value or the time for the base value. For example, if the time period is 12 hours, test data is observed at 1400, and base data is observed at 1800, data is accumulated from 0601 through 1800 hours. This has the effect that the tested time series and the base time series can have different observation times and different time intervals.

DISTRIBUTION <qflag> <duration> <significance_level> <base1> <base2>

abbr: DIS

This test is applied to incremental values. <base1> and <base2> are labels identifying time series at two nearby gages. The tested time series and the two base time series are divided into events. Events are separated by periods of zero values occurring simultaneously at all three gages. <duration> specifies the length of these zero-value periods. A chi-squared variable is used to compare the tested time series against both base time series. If the chi-squared probability is less than <significance_level> for both base time series, the quality flag will be set to <qflag> for all values in an event.

The chi-squared variable is computed by placing values in three classes and comparing the number of values in each class. The three classes are:

0
0 to 0.5 max_value
0.5 max_value to max_value,

where max_value is the maximum value, from all three time series, occurring in an event.

<significance_level> is the probability that the tested time series will be flagged as different when it is actually from the same distribution as the base time series. See Appendix B.

ESTIMATE <qflag> <method>

Specify estimation method. Estimates are made for values in DATA with flags specified in <qflag>. Values are not estimated for protected data.

<qflag> is one or more of

A = acceptable,
Q = questionable,
R = reject, and
M = missing

If there is not sufficient good data to make an estimate, no estimate is made; rejected values are then replaced by a value indicating missing data. Currently available methods are linear and missing.

LINEAR (<nmax> [<action>]) interpolates estimates for values in DATA with flags matching <qflag> using linear interpolation. Estimates will be made for up to <nmax> contiguous values. <qflag> characters must be Q, R, or M. Estimates are based on values with quality flags "better than" <qflag>. That is, replacement values for "rejected" data are based on"questionable" or "acceptable" data; replacement values for "questionable" data can only be based on "acceptable" data.

If more than <nmax> contiguous values are to be estimated, rejected values are replaced by missing values, and questionable values will be replaced by missing values unless <action> is ACCEPT.

MISSING sets data to a value indicating missing data for data with quality matching <qflag>.

If no estimation method is given, reject data will be replaced by values indicating missing data. Questionable data are not changed.

TEST <name> ...

Identify one or more previously-defined tests to be applied to time series specified on the following DATA instructions.

DATA <source pathname> ; <destination pathname>

Specify pathname for data to be screened.

Following the source pathname and separated from the source pathname by a semicolon (;), give the parts of the destination pathname (including file:) which differ from the source pathname.

If no parameters are given for the destination pathname, revised data will not be written to DSS.

Examples:

```
DATA DSSFIL:/TEST/LOCA/FLOW//1HOUR/RAW/;F=REV
DATA DSSFIL:B=LOCB F=RAW; F=REV

DATA RAWDSS:/TEST/LOCA/FLOW//1HOUR/OBS/; REVDSS:
DATA RAWDSS: B=LOCB; REVDSS:
```

END

Indicates end of a test definition or the end of input data.

(This page intentionally left blank)

Chapter 4

Viewing Program (DATVUE)

4.1 Introduction

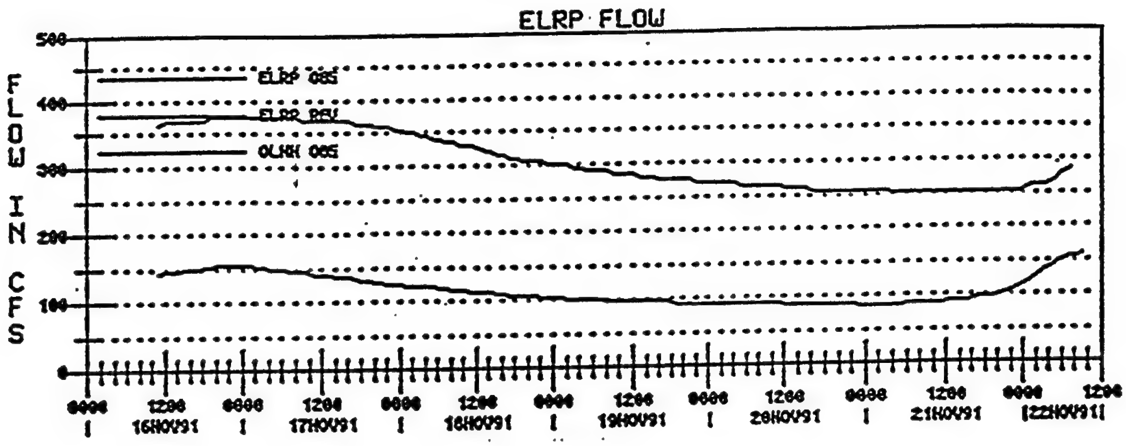
This Chapter describes the capabilities of DATVUE, the interactive program for viewing and editing data previously screened by DATCHK. The use of computer program DSPLAY to generate graphs for viewing within DATVUE, and the development and use of data-status and data-range tables are also described.

4.2 Overview

A primary function of DATVUE is to provide capability to easily and quickly review and edit data that has previously been rejected or categorized as questionable by DATCHK. Figure 4.1 shows a plot and scrollable table for a set of data such as could be displayed on a terminal screen by DATVUE. Data values in the "Actual" column are raw-data values; those in the "Revised" column can be edited by the analyst and are stored in the DSS file containing screened data. Codes in the table indicate values that were flagged by DATCHK (e.g., with a "Q" or "R"), and also show the tests that are failed (e.g., with a "RM" for Relative Magnitude). Editing options are discussed later in this Chapter.

DATVUE has an "auto" option with which the program will automatically produce a sequence of displays (such as that illustrated in Fig. 4.1) for each data record containing questionable or rejected data values. This option enables rapid review of this subset of data.

The plot shown in Fig. 4.1 was generated with the program DSPLAY, using the program's "save-plot" option. Plots can be set up to include data for nearby sites in addition to the site of interest. Because the plots are generated externally from DATCHK/DATVUE, the plots are not updated as data is edited. It is necessary to re-execute DSPLAY if revised plots are desired. Use of DSPLAY macros that employ the save-plot option is described in this Chapter.



17 /usr/orp/aleg/revdb.dss:/ALLEGHENY/ELRP/FLOW//1HOUR/REV/

Date	Time	Actual		Estimate	Revised			92 Q	0 R
21NOV91	2100	103.6	Q	—	103.6	a	RM		
21NOV91	2200	107.8	Q	—	107.8	a	RM		
21NOV91	2300	112.0	Q	—	112.0	a	RM		
21NOV91	2400	116.5	Q	—	116.5	a	RM		
22NOV91	0100	121.0	Q	—	121.0	a	RM		
22NOV91	0200	127.0	Q	—	127.0	a	RM		
22NOV91	0300	134.5	Q	—	134.5	a	RM		
22NOV91	0400	140.5		—	140.5	a			
22NOV91	0500	145.6		—	145.6	a			
22NOV91	0600	151.0		—	151.0	a			
22NOV91	0700	152.8		—	152.8	a			
22NOV91	0800	156.4		—	156.4	a			
22NOV91	0900	160.0		—	160.0	a			

Figure 4.1 Plot and Scrollable Table for Data Editing

In addition to the edit capabilities of DATVUE, the program has the following capabilities:

- generate a DSS test-criteria file from a text file

- manage a list of "gages-to-be-ignored"; data records for gages on this list will be set to "missing" by DATCHK

- generate a random-access data status file from a list of pathnames

- display the log file, data status files, and list of "gages-to-be-ignored"

4.3 Commands

These commands can be used when DATVUE is run as an interactive program.

AUTO [<number>]

Set review mode to automatic. The log file is checked for data records with questionable data, and these are displayed for editing. <number> is from the list of pathnames displayed by the DPATH command. If <number> is given, the associated time series is displayed whether or not it contains questionable data, then remaining time series with questionable data are displayed.

CONVERT <file>

Convert test criteria from text form to DSS form for use by DATCHK. <file> is the text file which is to be converted.

CONVERT.D <textfile> <randomfile>

Convert pathname list in text file into a random-access data-status file. The random-access file is updated by DATCHK and can be viewed using DATVUE's STATUS command.

CRITFILE <file>

Open a DSS test criteria file. This file is used to create a list of pathnames for data to be displayed in DATVUE.

DEVICE <device>

The DEVICE command is used to specify the type of terminal being used and the environment under which it will function. If no parameters are entered, no change will occur. The only valid graphics terminal types are the Tektronix and X-Windows types specified below.

If any other terminal type defined in the terminal definition file is used, the program will assume that it is a not a graphics terminal.

Options:

4107
XTERM
V7
VCOM

The XTERM device has the following parameters:

XPOSITION=	Upper left hand corner x-coordinate of graphics window. default=1
YPOSITION=	Upper left hand corner y-coordinate of graphics window. default=1
WIDTH=	Width of graphics window (0-1024). default=1000
HEIGHT=	Height of graphics window (0-1024). default=350
INVERT	Used to invert the background and foreground colors.
FONT	Used to set the font size. There are seven font sizes ranging from 1-7. The default size is set to 2.

Examples: DEVICE XTERM WIDTH=1000 HEIGHT=350 XP=1 YP=1 FONT=2 INVERT
DEV XT WI=400 INVERT

DISPLAY<number>
or
DISPLAY <pathname_parts>

Display DSS data for the pathname specified by <number>. <number> is from the list of pathnames displayed by the DPATH command.

If pathname parts are specified on the DISPLAY command, data will be displayed for the first pathname matching the specified pathname parts.

Examples:

di 17
di b=elrp

DPATH.option [<pathname_parts>]

Display a list of pathnames matching specified pathname parts from the test criteria file. If no pathname parts are specified, all pathnames in the criteria file will be listed.

Option Q - only display pathnames for records containing questionable or reject data.
The time-series data can be viewed using the DISPLAY command. After a time series has been reviewed its pathname is marked with a plus sign (+).

The number of questionable, reject, and missing data from the most recent DATCHK run are shown for each pathname. If the number of values in a category is greater than 99, double asterisks (**) are shown for that category.

FINISH

Exit from DATVUE.

GAGE

Display list of gages to be ignored.

GAGE.I <pathname> [<time window>]

Add this pathname to the list of gages to be ignored. An entry with a time window overlapping the specified time window will be replaced with the new time window.

GAGE.R <pathname> [<time window>]

Restore the gage identified by pathname to active status. All entries within the specified time window will be removed from the list of gages to be ignored.

On the GAGE command, <pathname> can be specified using the full pathname or using the index number from the pathname list displayed by the DPATH command. Once a pathname has been specified, subsequent entries will default to the previous pathname, and new pathnames can be entered by specifying new pathname parts.

The default time window is from beginning to end. Once a time window has been specified, that time window will be used as the default time window for subsequent entries.

GAGEFILE <file>

Open a gage list file. This file lists gages with data to be ignored; that is , reported values will be replaced with missing values for the specified time window. The default gage list file is "gagelist". Use this command to specify a different file.

HELP [<command>]

Display a list of DATVUE commands. If <command> is specified, display a description

Display a list of DATVUE commands. If <command> is specified, display a description of the command.

PLOTFILE <file>

Open a save-plot file created by DSPLAY. This file is the source of plots saved using the SPLOT command in DSPLAY.

STATFILE <file>

Open a data status file. The default file name is "statabl".

STATUS.option

Display the data-status file. Quality flags for the last 44 hours are displayed.

Options:

R - Display range categories
F - Copy data status to a text file
AF - Append data status to a text file

TIME <start data>, <time>, <end time>, <time>

Set time window for data to be displayed in DATVUE. This window will override the time window specified in the DSS criteria file.

VLOG.option

View contents of log written by DATCHK for <date>. The current date is used if no date is given. The last page of the log will be displayed first. Use + and - keys to page through the log file. Log entries for individual questionable values are not shown on the terminal screen. Use option D with option F or P to make a copy of the log showing data marked as questionable or reject.

Options:

D - write log entries for questionable data; use this option with F or P
P - pring log file
F - copy log to a text file

WHEN [<file>]

Search log files and display most recent times when DATCHK and DATVUE opened criteria files. If <file> is not named, the times for the most recently opened criteria file will be displayed.

4.4 Data Editing

When the DISPLAY command is used, a table of values appears on the screen. The actual data, estimated, and revised values are shown in the table.

Date	Time	Actual	Estimate	Revised
ddmmmyy	hhmm	nnnnnn f	nnnnnn f	nnnnnn f tf
ddmmmyy	hhmm	nnnnnn f	nnnnnn f	nnnnnn f tf
ddmmmyy	hhmm	nnnnnn f	nnnnnn f	nnnnnn f tf
ddmmmyy	hhmm	nnnnnn f	nnnnnn f	nnnnnn f tf
ddmmmyy	hhmm	nnnnnn f	nnnnnn f	nnnnnn f tf

Data in the column headed 'Actual' are the observed values from DSS, 'Estimated' are suggested values provided by DATCHK or DATVUE, and 'Revised' are the values which will be written to DSS to replace questionable values. The 'f' following the Actual value is a flag indicating data quality of the Actual value.

Data quality flags for actual values are 'R' for reject, 'Q' for questionable, and blank for missing and acceptable data.

Estimate flags indicate the method used for producing the estimate, such as 'I' for interpolation or 'M' for missing.

If a value is not questioned, the actual value will appear in the revised column by default, otherwise the estimated value appears in the revised column. Any value can be replaced with a new value whether it is questioned or not.

With this display the user can:

Scroll forward or backward (the cursor will be in the Revised column);
Press '?' to display editing commands;
" 'N' to find next questionable or missing value;
" 'B' to move back to previous questionable or missing value;
" 'T' to enter and move to a desired time;
" 'A' to accept actual value;
" 'E' to accept estimated value;
" 'M' to replace actual value with missing-data value; or
Enter a new value to replace actual value;
Press 'Q' to mark a value as questionable, then
" 'I' to interpolate values for all data marked with a 'Q';
" 'F' when finished with editing; or
" 'X' when finished with editing and exit from AUTO mode;
" 'R' to toggle repeat function; When Repeat is ON, keys 'A', 'E', 'M'. 'P'. 'Q', and 'U' will be repeated for all contiguous values with the same quality and test flags as the value on the current line.
" 'G' to add the current gage to the gagelist
" 'C' to display test criteria

Revision flag indicates how revised value was set:

blank - no revision

A - actual value
E - estimated value
I - interpolated value
M - missing value
Q - value is marked for interpolation
K - keyboard entry

The revision flag may have a second character indicating if the value is protected:

blank - value is not protected

P - value is protected
U - value will not be protected when changes are saved

Use these keys to change the protect flag:

> Press 'P' to set protection flag;
> Press 'U' to unprotect (clear the protection flag).

Following the replacement flags are 2-character test flags, tf, indicating the test which "questioned" the actual data. These flags are:

- AM - absolute magnitude test
- DM - duration-magnitude test
- RM - relative magnitude test
- CV - constant value test
- RC - rate-of-change test
- GL - gage was found in list of gages to be ignored
- NI - negative incremental value
- DS - distribution test

DATVUE uses key definitions from a terminal-definition file "coed.trm". DATVUE recognizes the definitions for:

- CURU - cursor up
- CURD - cursor down
- PGUP - page up
- PGDN - page down
- ENTR - enter (carriage return)
- BKSP - move cursor left, erase character
- HOME - move to top of list
- END - move to bottom of list

4.5 Graphical Display

Plots can be displayed simultaneously with tabulated data on Tektronix 4100-series terminals or in an Xterm Window. Plots are created by DSPLAY and saved in a file. Then DATVUE recalls the plot and displays it. With Tektronix emulation, the plot is displayed on the upper part of the screen, while tabulated data are shown on the bottom part. With an Xterm device, the plot is displayed in a separate window. When DSPLAY is executed, name the file where plots are to be saved with the SPLOT parameter:

```
dsplay splot=file
```

The following DSPLAY macro illustrates how to save plots. The files and pathnames will have to be modified for a particular application. Note that the plot is saved in a DSS file (option Z) using pathname parts A, B and C. Part D is set to /SCREEN-PLOT/. DATVUE uses

pathname parts A, B, and, C from the raw-data pathname to find the plot in the plot file. An example oriented to Xterm usage is included in Chapter 5.

```
MACRO SAVPLOT APART BPART CPART
SPLOT.Z PLOTFILE:/APART/BPART/CPART/SCREEN-PLOT/// start saving plot
LE.L OFF                                    turn off legend
DATE OFF                                    turn off date
FRAME OFF                                   turn off frame
SCR 100,920,480,750                         set plot dimensions
DEV SG=OFF                                  turn off Tektronix segments
US BPART CPART                              set user label
OP raw-data file                            open raw-data file
PA /APART/BPART/CPART//e/f/                 set raw-data pathname
OP revised-data file                        open revised-data file
PA /APART/BPART/CPART//e/f/                 set revised-data pathname
PL                                          plot data
QUIT                                        stop saving plot info
ENDMACRO
```

The expected range of values should be set. Otherwise a bad value might cause the scale to be so large that good data is compressed into a straight line on the plot.

```
YR.S 0,20
!RUN SAVPLOT GOES ACSOF STAFF
```

will save a plot of STAGE at location ACSOF. DATVUE will look for this plot under the pathname /GOES/ACSOF/STAGE/SCREEN-PLOT///.

To use the plots in DATVUE, set the plot file using the PLOTFILE parameter:

```
datvue plotfile=file
```

or use DATVUE's PLOTFILE command.

If a plot is not found when DATVUE is asked to display data, the whole screen is used for tabulated data.

Example files and macros for using DSPLAY are shown in Chapter 5.

4.6 Data Status and Range

DATVUE has capability to display a data-status table (with the command STATUS), and a data-range table (with the command STATUS.R). Information in these tables is maintained by DATCHK when data is screened. The status information is shown for the most recent 44 hours. The following is an illustration of a data-status table:

```
           Data Status for Smith River

               *** Precipitation Gages ***
Location  Parameter              12Jan90          |          13Jan90
ABC   PRECIP-INC        ---- ---- ---- ---- ---- R--- ---- ---- --..
DEF   PRECIP-INC        ---- ---- MM-- ---- ---- ---Q QQQQ Q--- ---.
GHI   PRECIP-INC        ---- ---- ---- ---- ---- ---- ---- ---- --..
JKL   PRECIP-INC        .... .... .... .... .... .... .... .... ....

            ***  Flow Gages ***
MNO             Flow    ---- ---- ---- ---- ---- ---- ---- ---- --..
```

Data-status symbols are:

. no data for this time, or data not screened
- data is acceptable
M data is missing
Q questionable value
R reject value

The data status table shows 44 one-hour time slots. If more than one value occurs in a time slot, the quality flag is for the value with the poorest quality and is shown in lower case.

The following is an illustration of a data-range table:

```
           Data Status for Smith River

                *** Precipitation Gages ***
Location  Parameter             12JAN90          |            13JAN90
ABC   PRECIP-INC       ---- ---- ---- ---- ---- 3--- ---- ---- --..
DEF   PRECIP-INC       ---- ---- mm-- ---- ---- ---1 0122 1--- ---.
GHI   PRECIP-INC       ---- ---- ---- ---- ---- ---- ---- ---- --..
JKL   PRECIP-INC       .... .... .... .... .... .... .... .... ....

              *** Flow Gages ***
MNO    FLOW            ---- ---- ---- ---- ---- ---- ---- ---- --..
```

Data-range symbols are:

.	no data for this time
-	data value is between zero and the first range limit
1	data value is between the first and second range limits
2	data value is between the second and third range limits
3	data value is above the third range limit
m	data value is missing

The range limits are specified with the DATCHK "RANGE" command.

To produce the above tables, a text file containing a list of pathnames for the tables is required. The following is an example of such a text file:

```
                Data Status for Smith River

        *** Precipitation Gages ***
    /SMITH/ABC/PRECIP-INC//1HOUR/OBS/
    /SMITH/DEF/PRECIP-INC//1HOUR/OBS/
    /SMITH/GHI/PRECIP-INC//1HOUR/OBS/
    /SMITH/JKL/PRECIP-INC//1HOUR/OBS/

        *** Flow Gages ***
    /SMITH/MNO/FLOW//1HOUR/OBS/
```

A random access file (for use by DATCHK) is generated from the text file using the CONVERT.D command in DATVUE. DATCHK will update the random-access file as it screens the data.

Chapter 5

Program Usage

5.1 Introduction

To setup a data-screening procedure involves preparation of a number of files, as described previously in this Manual. This chapter summarizes information on file purpose and usage, and provides guidelines for implementing data-screening procedures. Figure 5.1 shows data files associated with the screening process. The dashed circles are the basic files prepared by the user.

5.2 DATCHK Files and Parameters

File names and parameters can be set on the "execution line" that calls for program execution. The user can obtain a list of files and parameters, and their default values, by typing.

datchk \? (Note: a "\" is not used on the Harris minicomputer)

The list is as follows:

```
DATCHK:   1.1.2    23 Jan 95
   UNIT    KEYWORD    *AVREV     **MAX      DEFAULT
    6      OUTPUT     0           83        /dev/tty
  NOP      INPUT      I           83
  NOP      CRITFILE   C           83
  NOP      GAGEFILE   G           83        gagelist
  NOP      STATFILE   S           83
  NOP      DATE       D            9        current date
  NOP      TIME       T            4        2400
   30      SCRATCH    SC          83        SCRATCH.011
  NOP      FUNFILE    F           83        genfun

* ABREV - SHORTEST ABBREVIATION ALLOWED FOR KEYWORD
** MAX - MAXIMUM # OF CHARACTERS FOR FILENAME (OR STRING)
```

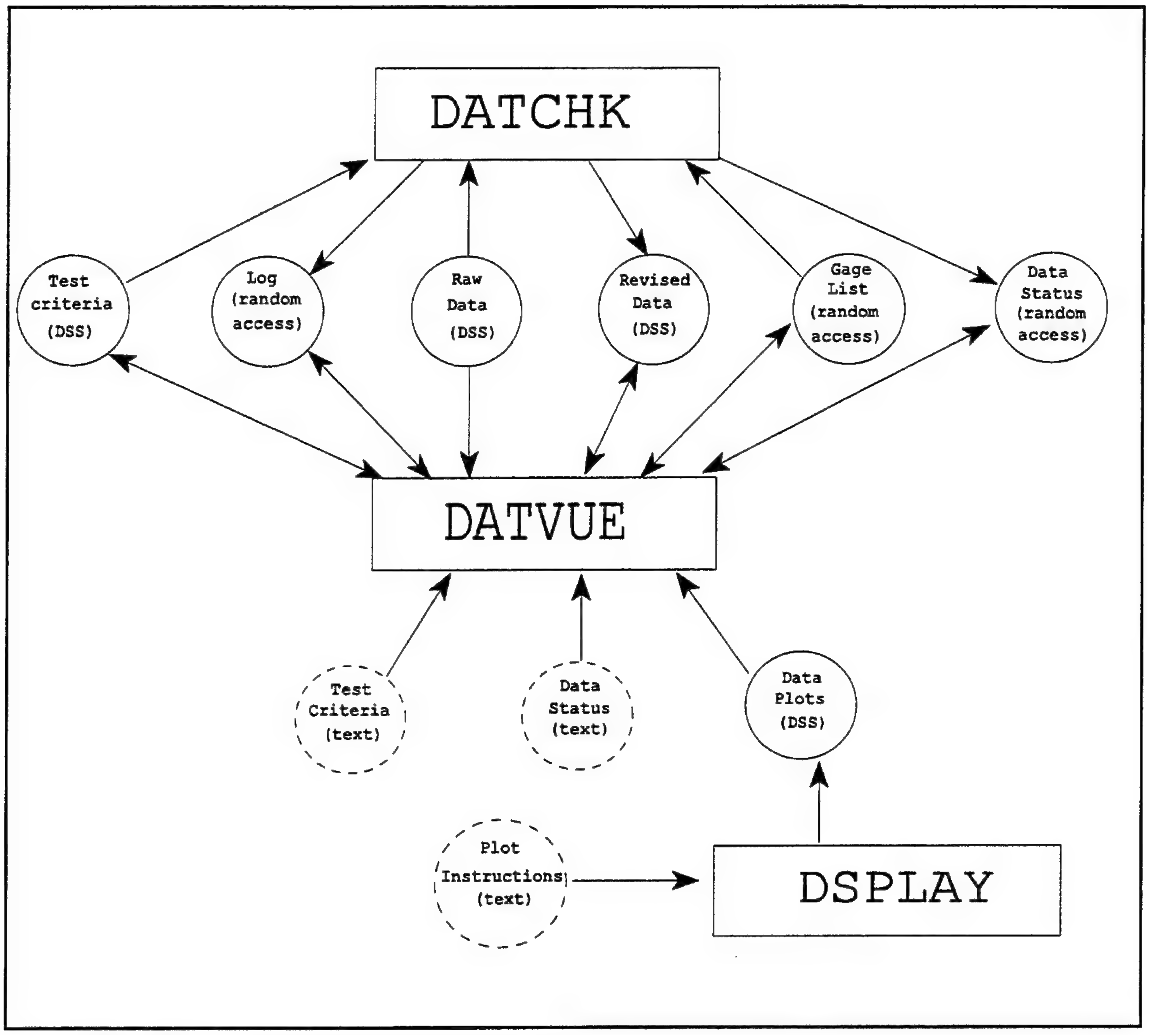


Figure 5.1 Files

OUTPUT is a file that contains, as a minimum, a summary of the number of reject, questionable and missing values found by application of screening tests. Higher levels of output can be obtained by setting the output level in the INPUT file, commands for which will be presented subsequently.

INPUT is a file that can optionally be used to restrict the screening process to a subset of the pathnames contained in the DSS criteria file, CRITFILE. Commands that can be used in the INPUT file are described in the next Section.

CRITFILE is a DSS criteria file that is generated from a text criteria file using DATVUE. The name of the file is specified with the DATCHK "CRITFILE" command.

GAGEFILE is a random access file created and modified with the DATVUE "GAGE" command. The file is a list of gages to be ignored. The default name for this file is "gagelist". That is, if such a file exists, it need not be named on the execution line for DATCHK.

STATFILE is a random-access file created with the DATVUE "CONVERT.D" command from a text file containing a list of pathnames. The names of both files are parameters of the "CONVERT.D" command.

DATE and TIME can be optionally specified to set values for Refdate and Reftime. These are used in conjunction with the DATCHK "TIME" command.

FUNFILE is an optional "function" file that contains function references (see documentation for PREAD capabilities).

5.3 DATCHK Input-file Commands

These commands are used in the file named by DATCHK's execution-line parameter INPUT. When an INPUT file is used, DATCHK will screen only those time series named in the INPUT file. These commands can also be used to override some of the parameters in the DSS criteria file.

CRITFILE <file>

Open a DSS test criteria file. This command can be used to open alternative criteria files.

DATA <source pathname>

Specify pathname for data to be screened. If <source pathname> is ALL, then DATCHK will screen time series for all records in the DSS criteria file.

GAGEFILE <file>

Open a gage list file. Data for pathnames and times listed in the gage list file will be rejected. (Use DATVUE to create and edit the gage list file.)

OUTPUT <level>

Controls amount of output written by DATCHK. Output <level>s are:

1 - minimum output, summary of number of reject, questionable, and missing values found;

2 - level 1 plus shows pathname and data quality flags after all tests have been completed;

3 - level 2 plus shows test parameters and quality flags for tests which find questionable data;

4 - level 3 plus test parameters for tests which did not find questionable data;

5 - level 4 plus quality flags for tests which did not find questionable data.

Default is output level 1.

PUT <label> = <pathname>

Write the time series associated with <label> to DSS using <pathname>. <pathname> includes the DSS file name.

STATFILE <file>

Open a data-status file. Data-status and range flags will be written to this file. Use DATVUE to view the data-status file.

TIME <start date> <start time> <end date> <end time>
or
TIME T-nX [T-nX]
or
TIME -nX [-nX]

Specify update time window. See description of TIME instruction in Section 3.4 for details. This instruction will override the time window saved in the DSS criteria file.

TRACE ON | OFF

Turn trace on or off. When trace is on, the time series will be written to the output file after each operation that changes the time series, and after each test that detects questionable data.

5.4 DATVUE Files

File names and parameters for DATVUE can be listed by typing

datvue \?

The list is as follows:

```
DATVUE: 1.0.3  01 Dec 94
   UNIT       KEYWORD     *ABREV      **MAX       DEFAULT
      7       INPUT       I            83         /dev/tty
      6       OUTPUT      O            83         /dev/tty
    NOP       CTEXT       C            83
    NOP       TRMDEF      T            83         /usr/hec/sup/coed.trm
    NOP       FUNFILE     F            83         genfun
    NOP       MACFILE     M            83         dvumac
    NOP       SCNFILE     S            83         dvuscn
     35       HELPFILE    H            83         /usr/hec/sup/datvue.hlp
    NOP       PLOTFILE    P            83
     30       SCRATCH     SCR          83         SCRATCH.001

   * ABREV - SHORTEST ABBREVIATION ALLOWED FOR KEYWORD
   ** MAX - MAXIMUM # OF CHARACTERS FOR FILENAME (OR STRING)
```

INPUT is normally from the keyboard, although it is possible to execute DATVUE with an input file. OUTPUT is normally directed to the terminal screen.

CTEXT is the name of the text criteria file. To generate a DSS criteria file from the text file, DATVUE can be executed as follows:

datvue ctext=filename

DATVUE generates the DSS file, then stops.

TRMDEF is the terminal definition file. This file defines the behavior of the keyboard to facilitate editing procedures. The default file "coed.trm" is normally used.

FUNFILE, MACFILE and SCNFILE are optional files containing function references, macros and screens, respectively. These are associated with PREAD capabilities.

HELPFILE contains command descriptions that are accessed with the "help" command.

HELPFILE contains command descriptions that are accessed with the "help" command. The default file "datvue.hlp" is normally used.

PLOTFILE is the file containing DSPLAY-generated "save plots". DATVUE will access this file to obtain plots to display with a scrollable table for data editing.

5.5 Implementation of Data-Screening Procedures

The following are guidelines and illustrations for implementation of data-screening procedures.

1. Determine how data types will be grouped for screening purposes. In the illustration that follows, separate files are used for discharge data and precipitation data.

2. Develop (text) criteria files. Fig. 5.2 illustrates a criteria file for discharge data, and Fig. 5.3 for precipitation data. Generally tests (criteria) for variables such as discharge or stage must be site specific, whereas with precipitation data substantial use can be made of global tests.

3. Generate DSS criteria files from the text files with DATVUE. In this illustration, this can be accomplished with the following two executions of DATVUE:

datvue ctext=flocrit.tx

datvue ctext=pptcrit.tx

This results in the creation of the files "flocrit.dss" and "pptcrit.dss", the names for which were specified with the CRITFILE command in the text files.

4. Develop input and macro files for use in generating plots with DSPLAY. Such files are illustrated in Figs. 5.4-5.7. The macros are set-up to include data for nearby gages. Precipitation data is plotted in both incremental and cumulative forms.

5. Develop text files that contain pathnames for data-status and data-range tables. Such files are illustrated in Figs. 5.8 and 5.9.

6. Generate random access status files from the text files using the "CONVERT.D" command in DATVUE. For this illustration, the resulting files are labeled "stat_flo" and "stat_ppt".

7. Create a job file for executing DATCHK and DSPLAY. The following illustrates contents of such file:

```
rm svpl.dss
datchk c=flocrit.dss s=stat_flo o=datchk_flo_out g=glist_flo
dsplay -C i=dsplay_in_flo m=svplmac_flo o=svpl_flo_out term=/dev/null
```

```
datchk c=pptcrit.dss s=stat_ppt o=datchk_ppt_out g=glist_ppt
dsplay -C i=dsplay_in_ppt m=svplmac_ppt o=svpl_ppt_out term+/dev/null
```

The job file would normally be executed automatically as part of the data acquisition process.

8. The preceding step completes the basic file-preparation requirements. When DATCHK executes, it will access raw data from the file "mastdb.dss" and store screened data in the file "revdb.dss". DSPLAY will generate plots and store them in the file "svpl.dss". Data review and editing can then be performed with DATVUE. To facilitate the start-up of DATVUE, a bootstrap macro can be used. Such a macro is illustrated in Fig. 5.10, which displays the file "dvumac".

The bootstrap macro automatically opens the plot file "svpl.dss", specifies the plot device, executes the "when" command to show when DATCHK and DATVUE were last executed, and enables the user to run the "flo" or "ppt" macro. These macros open the appropriate criteria, data-status, and faulty-gage files; then display names of records containing questionable data. At this point, the command "auto" can be entered to facilitate data editing.

Following completion of editing for one of the variables, editing of data for the other variable could be initiated by execution of the appropriate macro (e.g., "!r ppt" or "!r flo").

5.6 Screening Period-of-Record Data

Although DATCHK/DATVUE were designed to screen real-time data, they can also be used to screen period-of-record data. When doing period-of-record analysis, DATCHK/DATVUE partition the time series according to the array size declared in the programs.

In DATCHK, the partitions overlap by the time window extensions set on the CONTEXT instruction. DATCHK screens the data in each partition separately. All tests are applied to one partition, then data is read from DSS for the next partition.

In DATVUE, the partitions overlap by approximately 50 values. When a key, such as B, N, Page Up, Page Down, Up Arrow, or Down Arrow, is pressed and the next value is outside the current partition, DATVUE reads data for the new partition. Pressing Home or End twice moves the cursor to the beginning or end of the current partition. To display data for a particular date, press T and enter the date. If the date is not in the current partition, DATVUE defines a new partition centered around the requested date. When DATVUE moves to a new partition it displays the messages: 'Reading data from DSS ...' and 'Updating arrays...'.

```
*    ***
*    ***  Textfile containing screening criteria for ALLEGHENY BASIN flow gages
*    ***  last updated:  22 JUL 1991
*    ***
CRITFILE flocrit.dss
*
TIME T-7D T
CONTEXT START -2D
*
*    ***    BRFP    ***
*
* Absolute max. criteria based on estimated Q's for PMF and 10% events
CRITERIA ABS R  0  65000
CRITERIA ABS Q  0 11500
CRITERIA CONST Q 2D 0 0 3
*  Rate-of-change criteria based on June 89 event + 25%
CRITERIA RATE Q -2950 +4800
ESTIMATE RM LINEAR 3
DATE /usr/orp/mastdb.dss:/ALLEGHENY/BRFP/FLOW//1HOUR/OBS/; /usr/orp/aleg/revdb.dss: F=REV
*
*    ***    BRKP    ***
*
*  Absolute max. criteria based on estimated Q's for PMF and 10% events
CRITERIA ABS R   0 120000
CRITERIA ABS Q   0 14500
CRITERIA CONST Q 2D 0 0 3
*  Rate-of-change criteria based on June 89 event + 25%
CRITERIA RATE Q -2950 +4800
ASSIGN.R SNCP = /usr/orp/mastdb.dss:/ALLEGHENY/SNCP/FLOW//1HOUR/OBS/ O  75000
COMPUTE TSNCP = TSHIFT (SNCP, -8H)
*  Rel. mag. based on Q being > 25% and < 95% of TSNCP
CRITERIA REL Q 0.25*TSNCP  0.95*TSNCP
ESTIMATE RM LIENAR 3
DATA /usr/orp/mastdb.dss:/ALLEGHENY/BRKP/FLOW//1HOUR/OBS/; /usr/orp/aleg/revdb.dss: F=REV

       * * * *  Criteria for remaining stations follow here * * * *

END
```

Figure 5.2 File "flocrit.tx"

```
* ***
* ***  Textfile containing screening criteria for ALLEGHENY BASIN precip. gages
* ***  last updated:  9 August 1991
* ***
CRITFILE pptcrit.dss
*
TIME T-7D T
CONTEXT START -24H
*
DEFINE PPT1
  CRITERIA  ABS  Q   0   1.7
  CRITERIA  ABS  R   0  14.4
*
  CRITERIA  DUR  Q   0   2.4    3H
  CRITERIA  DUR  Q   0   2.8    6H
  CRITERIA  DUR  Q   0   3.1   12H
  CRITERIA  DUR  Q   0   3.3   24H
*
  CRITERIA  DUR  R   0  20.3    3H
  CRITERIA  DUR  R   0  25.9    6H
  CRITERIA  DUR  R   0  35.5   12H
  CRITERIA  DUR  R   0  41.5   24H
*
  ESTIMATE  R  MISSING
  PRECISION 2
*
* Range boundaries based on 2-yr, 5-yr and 10-yr recurrence intervals
*
  RANGE 1.1   1.5   1.7
END
*
*
* ** ALMN **
  TEST PPT1
  AS.A SNGP = /usr/orp/mastdb.dss:/ALLEGHENY/SNGP/PRECIP-INC//1HOUR/OBS/
  AS.A FNDN = /usr/orp/mastdb.dss:/ALLEGHENY/FNDN/PRECIP-INC//1HOUR/OBS/
  CRITERIA DIS Q 6H  .02 SNGP FNDN
  CRITERIA REL  Q  .2*SHGP-.5  3*SNGP+.5  ACCUM  6D
  DATA /usr/orp/mastdb.dss:/ALLEGHENY/ALMN/PRECIP-INC//1HOUR/OBS/;/usr/orp/aleg/revdb.dss:F=REV
*
* ** BGBP **
  TEST PPT1
  AS.A GVLP = /usr/orp/mastdb.dss:/BEAVER/GVLP/PRECIP-INC//1HOUR/OBS/
  AS.A MERP = /usr/orp/mastdb.dss:/BEAVER/MERP/PRECIP-INC//1HOUR/OBS/
  CRITERIA DIS Q 6H  .02 GVLP MERP
  CRITERIA REL Q  .2*GVLP-.5  3*GVLP+.5  ACCUM  6D
  DATA /usr/orp/mastdb.dss:/BEAVER/BGBP/PRECIP-INC//1HOUR/OBS/;  /usr/orp/aleg/revdb.dss:F=REV
*
            * * * *  Criteria for remaining stations follow here  * * * *

END
```

Figure 5.3 File "pptcrit.tx"

```
/usr/orp/mastdb.dss
! -ECHO
TIME T-7D  T
DATE OFF
FRAME OFF
DGRID GRID=MAJOR STYLE=DOTTED
DGRID GRID=MINOR STYLE=NONE
TYPE 1 1 1 1 1
!*SCREEN 100 920 480 750
SCREEN 100 920 210 740
YR.S 0, 500
!*DEVICE BATCH SF=OFF 4107
DEVICE BATCH XTERM GIN=OFF WI=1023 HE=350 XP=0
!*LE.L 4 1 4 110 300 620 740
LE.L 4 1 4 110 1000 5 145
LE.XT B
LE.A 1 OBS
LE.A 2 REV
LE.A 3 OBS
LE.A 4 OBS
DL 1 COLOR=RED
DL 2 COLOR=BLUE
DL 3 COLOR=GREEN
DL 4 COLOR=CYAN
!RUN SAVPLOT BRFP FLOW
!RUN SAVPLOT1 BRKP FLOW SNCP
!RUN SAVPLOT BVLP FLOW
!RUN SAVPLOT CLAP FLOW
!RUN SAVPLOT CNOP FLOW
!RUN SAVPLOT1 COKP FLOW RDYP
!RUN SAVPRES CONP FLOW-RES.IN
!RUN SAVPRES CRCP FLOW-RES.IN
!RUN SAVPLOT CROP FLOW
!RUN SAVPLOT YVLP FLOR

 * * * *  Lines for remaining stations follow here  * * * *

FINISH
```

Figure 5.4 File "dsplay_in_flo"

```
MACRO SAVPLOT BPART1 CPART1
SPLOT.Z svpl.dss:/ALLEGHENY/BPART1/CPART1/SCREEN-PLOT///
US BPART1 CPART1
OPEN ../revdb.dss
PATH/usr/orp/mastdb.dss:/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/OBS/
PATH ../revdb.dss:/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/REV/
PLOT
QUIT
ENDMACRO

MACRO SAVPLOT1 BPART1 CPART1 BPART2
SPLOT.Z svpl.dss:/ALLEGHENY/BPART1/CPART1/SCREEN-PLOT///
US BPART1 CPART1
OPEN ../revdb.dss
PATH /usr/orp/mastdb.dss;/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/OBS/
PATH ../revbd.dss:/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/REV/
PATH /usr/orp/mastdb.dss:/ALLEGHENY/BPART2/CPART1//1HOUR/OBS/
PLOT
QUIT
ENDMACRO

MACRO SAVPLOT2 BPART1 CPART1 BPART2 BPART3
SPLOT.Z svpl.dss:/ALLEGHENY/BPART1/CPART2/SCREEN-PLOT///
US BPART1 CPART1
OPEN ../revdb.dss
PATH /usr/orp/mastdb.dss:/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/OBS
PATH ../revdb.dss:/ALLEGHENY/BPART1/CPART1/01JAN1990/1HOUR/REV/
PATH /usr/orp/mastdb.dss:/ALLEGHENY/BPART2/CPART1//1HOUR/OBS/
PATH /usr/orp/mastdb.dss:/ALLEGHENY/BPART3/CPART1//1HOUR/OBS/
PLOT
QUIT
ENDMACRO

MACRO SAVPRES BPART CPART
SPLOT.Z svpl.dss:/ALLEGHENY/BPART/CPART/SCREEN-PLOT///
US BPART CPART
OPEN ../revdb.dss
PATH /usr/orp/mastdb.dss:/ALLEGHENY/BPART/CPART/01JAN1990/1HOUR/COMP/
PATH../revdb.dss:/ALLEGHENY/BPART/CPART/01JAN1990/1HOUR/REV/
PLOT
QUIT
ENDMACRO
```

Figure 5.5 File "svplmac_flo"

```
/usr/orp/mastdb.dss
!-ECHO
TIME T-7D, T
DATE OFF
FRAME OFF
!*SCR 100,920,480,750
SCREEN 100 920 210 740
DGRID GRID=MAJOR STYLE=DOTTED
DGRID GRID=MINOR STYLE=NONE
!*LE.L 5,1,4,110,300,620,740
LE.L 4 1 4 110 1000 5 145
LE.XT B
LE.A 1 RAW
LE.A 2 REV
LE.A 3 RAW CUM
LE.A 4 RAW CUM
LE.A 5 RAW CUM
DL 1 COLOR=RED
DL 2 COLOR=BLUE
DL 3 COLOR=RED
DL 4 COLOR=GREEN
DL 5 COLOR=CYAN
TYPE 1 1 1 1 1
!*DEVICE BATCH SG=OFF 4107
DEVICE BATCH XTERM GIN=OFF WI=1023 HE=350 XP=0
!* SH 0,0,-1,-1,-1
YR.S 0, 0.3
!R SAVPLOT ALLEGHENY ALMN ALMN ALLEGHENY SNGP ALLEGHENY FNDN
!R SAVPLOT BEAVER BGBP BGBP BEAVER GVLP BEAVER MERP
!R SAVPLOT ALLEGHENY BRFP BRFP ALLEGHENY OLNN ALLEGHENY SLMN
!R SAVPLOT ALLEGHENY BRNP BRNP ALLEGHENY LYNP ALLEGHENY KZOP

         * * * *  Lines for remaining stations follow here  * * * *

FIN
```

Figure 5.6 File "dsplay_in_ppt"

```
MACRO SAVPLOT APART1 BPART1 BPART2 APARTX BPARTX APARTY BPARTY
SPLOT.Z svpl.dss:/APART1/BPART1/PRECIP-INC/SCREEN-PLOT///
USER BPART1 PRECIP-INC
PATH /APART1/BPART1/PRECIP-INC/01JAN1990/1HOUR/OBS/
PATH../revdb.dss:/ALLEGHENY/BPART2/PRECIP-INC/01JAN1990/1HOUR/REV/
OPEN/usr/orp/mastdb.dss
NOR.A ON
PATH /APART1/BPART1/PRECIP-INC/01JAN1990/1HOUR/OBS/
PATH /APARTX/BPARTX/PRECIP-INC/01JAN1990/1HOUR/OBS/
PATH /APARTY/BPARTY/PRECIP-INC/01JAN1990/1HOUR/OBS/
NOR.A OFF
PLOT
QUIT
ENDMACRO
```

Figure 5.7 File "svplmac_ppt"

```
                    * * * HEADWATER STREAMFLOW GAGES  * * *

/ALLEGHENY/BRFP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/BRKP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/BVLP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/DUJP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/ECMP/FLOW/01OCT1987/1HOUR/OBS/

                     * * * DOWNSTREAM STREAMFLOW GAGES * * *

/ALLEGHENY/CLAP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/COKP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/ELRP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/FRKP/FLOW/01OCT1987/1HOUR/OBS/
/ALLEGHENY/GRCP/FLOW/01OCT1987/1HOUR/OBS/

                    * * *  RESERVOIR OUTFLOW  * * *

/ALLEGHENY/TNOP/FLOW/01FEB1988/1HOUR/OBS/
/ALLEGHENY/UCOP/FLOW/01FEB1988/1HOUR/OBS/
/ALLEGHENY/CNOP/FLOW/01FEB1988/1HOUR/OBS/
/ALLEGHENY/CROP/FLOW/01FEB1988/1HOUR/OBS/

                    * * *  RESERVOIR INFLOW  * * *

/ALLEGHEHY/CONP/FLOW-RES.IN/01JUL1988/1HOUR/COMP/
/ALLEGHENY/CRCP/FLOW-RES.IN/01JUL1988/1HOUR/COMP/
/ALLEGHENY/EBRP/FLOW-RES.IN/01JUL1988/1HOUR/COMP/
/ALLEGHENY/LOYP/FLOW-RES.IN/01JUL1988/1HOUR/COMP/
```

Figure 5.8 File "databl_flo"

```
            * * *  PRECIPITATION GAGES  * * *

/ALLEGHENY/ALMN/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/BRFP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/BRKP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/BRNP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/BVLP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/CLAP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/COKP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/CONP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/CRCP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/CSPP/PRECIP-INC//1HOUR/OBS/
/ALLEGHENY/CTLP/PRECIP-INC//1HOUR/OBS/
/BEAVER/BTLP/PRECIP-INC//1HOUR/OBS/
/BEAVER/GVLP/PRECIP-INC//1HOUR/OBS/
/BEAVER/MERP/PRECIP-INC//1HOUR/OBS/
/BEAVER/SPPP/PRECIP-INC//1HOUR/OBS/
/MONONGAHELA/CHRP/PRECIP-INC//1HOUR/OBS/
/MONONGAHELA/CLLP/PRECIP-INC//1HOUR/OBS/
/MONONGAHELA/DGLP/PRECIP-INC//1HOUR/OBS/
/MONONGAHELA/WHBP/PRECIP-INC//1HOUR/OBS/
/MONONGAHELA/WILP/PRECIP-INC//1HOUR/OBS/
/OHIO/EMSP/PRECIP-INC//1HOUR/OBS/
/OHIO/FDMP/PRECIP-INC//1HOUR/OBS/
```

Figure 5.9 File "databl_ppt"

```
BOOTSTRAP
!RUN START

MACRO START
PLOTF svpl.dss
DEV XTERM  WIDTH=1023  HE=350  XP=1  YP=1
WHEN
!PRINT
!PRINT Enter "flo" to view flow data, or "ppt" to view precipitation data:
!KBLINE #
!RUN ^ #
ENDMACRO

MACRO FLO
CRITF flocrit.dss
STATF stat_flo
GAGEF glist_flo
DP.Q
ENDMACRO

MACRO PPT
CRITF pptcrit.dss
STATF stat_ppt
GAGEF glist_ppt
DP.Q
ENDMACRO
```

Figure 5.10 File "dvumac"

References

Hydrologic Engineering Center, 1992. DATCHK and DATVUE - Data Screening Software - Programmer's Manual (draft). U.S. Army Corps of Engineers, Davis California.

Hydrologic Engineering Center, 1994. HEC-DSS - User's Guide and Utility Program Manuals. U.S. Army Corps of Engineers, Davis, California.

Appendix A

Quality Flags

Quality Flags

Data quality flags are written to DSS with revised data. A flag consists of 4 bytes (32 bits), defined as shown below. DATVUE reads the flags when setting up the scrollable table for data editing.

1: Screened - set when original data has been screened

Quality of original data:

2: Okay
3: Missing
4: Questionable
5: Reject

6-7: Range of current data - an integer in [0,3]
8: Current value is different from original value

9-11: When set current value - an integer in [0,7]
- 0 original value, no revision
- 1 DATCHK
- 2 DATVUE
- 3 manual entry in DATVUE
- 4 original value accepted in DATVUE

12-15: Replacement method - an integer in [0, 15]
- 0 no revision
- 1 linear interpolation
- 2 manual change
- 3 replace with missing value

Tests Failed:

16: absolute magnitude
17: constant value
18: rate-of-change
19: relative magnitude
20: duration-magnitude
21: negative incremental value
22: not defined
23: gage list
24: not defined
25: user-defined test

26: distribution test

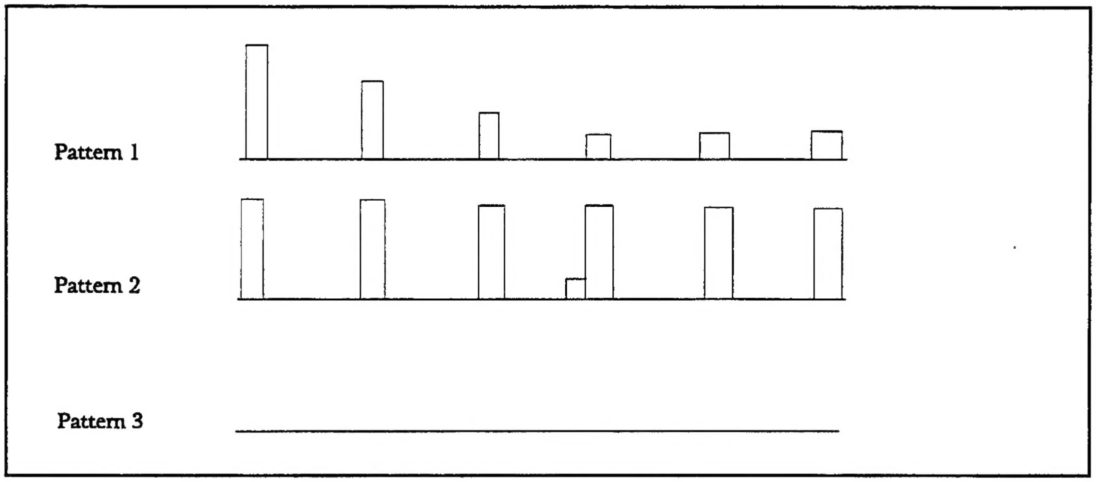
27-31: reserved for future use

32: protect value from being automatically changed

Appendix B

Distribution Test

Consider these rainfall patterns:



These are examples of patterns that may be questionable. Pattern 1 might occur when a rainfall collector is clogged with leaves, insects, or other debris. The rain water is sitting in the collector and slowly dripping through the clogged orifice. Pattern 2 might occur because of electronic failure in the recording or transmission devices. Pattern 3 is not obviously wrong unless rainfall at other gages is considered.

The distribution test was devised to detect patterns 1 and 2. It can also detect pattern 3.

The distribution test assumes the following:

1. Rainfall occurring at test gage and nearby gages are somehow related, and the time-distibutions of rainfall intensities are similar.

2. The maximum rainfall is approximately the same at all three gages.

3. Storms are separated by dry periods that equal or exceed a user-specified duration.

4. Rainfall-intensity distributions can be compared using a chi-squared probability distribution.

Procedure

The time series is divided into storms. A storm begins when rain begins to fall at any one of the three gages and ends when rain quits falling at all of the gages. Since there might be a lull in a storm, there is a further condition that the dry period between storms is at least 6 hours. Any rain falling within 6 hours of the end of a storm is considered to be a continuation of the storm. The storm is extended to include the new rainfall.

Using this definition of a storm allows the storm to travel through a region; i.e., rain does not have to be falling simultaneously at all three gages. However it is assumed that the storm is general in nature, so if rain occurs at two gages, it will also occur at the third.

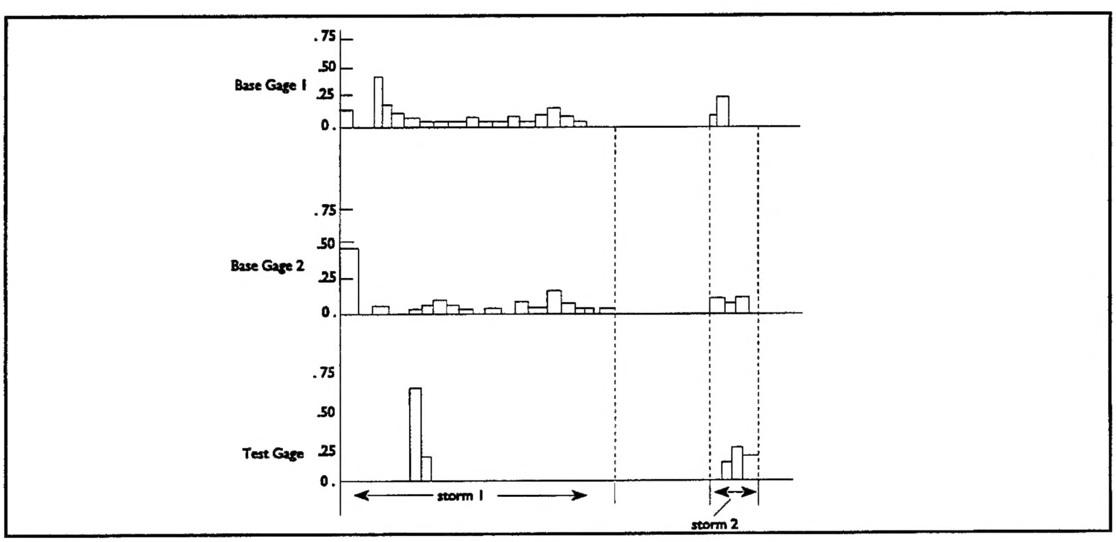
Once a storm is isolated the rainfall intensities are placed in three categories: no rain, 0 to .5 max, and .5 max to max, where max is the maximum rainfall occurring at all three gages.

A chi-square variable is computed using the number of values in the three categories for each pair of gages. If the probability of this variable is less than a threshold probabilitity, the two rainfall patterns are assumed to come from different distributions.

A rainfall pattern is questioned when

1) the patterns at the two base gages come from the same distribution, and

2) the questioned pattern is from a different distribution than both base gages.



Consider the rainfall distributions shown in the preceding figure. There are two storms shown, one lasting for 21 hours and another lasting 3 hours. For the larger storm, the maximum rainfall is .67 inches, which occurs at the test gage. The three rainfall categories and the number of values for each gage are:

Category	Base 1	Base 2	Test Gage
0	3	5	19
0-.33	17	15	1
.34-.67	1	1	1

Comparing the two base gages, the probability that their rainfall patterns are from the same distribution is 0.73.

The probability that the test gage has the same distribution as base gage 1 is 0.0000024; for base gage 2 the probability is 0.000037. Since these probabilities are very low, it is concluded that the rainfall recorded at the test gage is questionable for this storm. All 21 values are flagged.